

Kristen O'Grady

A DOQ TEST PROJECT: COLLECTING DATA TO IMPROVE TIGER

Abstract

DOQ is an acronym becoming quite well known to the U.S. Bureau of the Census, Geography Division research team investigating data collection methodologies with the goal of improving the positional accuracy of the Master Address File/Topologically Integrated Geographic Encoding and Referencing (MAF/TIGER®) system. The team is using Environmental Systems Research Institute's (ESRI) ArcView®¹ Geographic Information System (GIS) software as the vehicle for interpreting Digital Orthophoto Quadrangles (DOQ) and collecting data. ArcView offers several required functions, including the ability to easily reformat TIGER/Line '97 and DOQ data into data structures readable by ArcView GIS tools that aid photographic interpretation. This paper describes the DOQ test project, discusses the data collection process, and identifies successes and stumbling blocks.

Introduction

The Geospatial Research and Standards Staff (GRaSS) of the Census Bureau's Geography Division is currently investigating methods to improve the TIGER data base portion of the MAF/TIGER system. Updating the data base by adding new features and spatially enhancing existing features are two components of TIGER improvement. The GRaSS has underway two test projects to research these components. Updating the TIGER data base is being tested solely by collecting data with the use of GPS technology in the field. Improving the positional accuracy of and spatially enhancing TIGER is being additionally tested by collecting data in-house using DOQs. Both data collection techniques will be evaluated to determine which is a more efficient and feasible method of data collection to be considered for use on a nationwide basis.

DOQs are commonly used as source data for collecting digital information and in many other GIS applications. However, the DOQ test project is the first of its kind at the Census Bureau and is viewed as a non-traditional means of data collection. The GRaSS has completed the first of two phases of the DOQ test project and has encountered both successes and obstacles. As one of the two major data collectors for the DOQ test project, the author's participation has been challenging given limited prior experience with DOQs and photographic interpretation. One result of the author's novice status has been a somewhat unique perspective concerning GIS tools that might better facilitate data collection and improve photographic interpretation for future DOQ projects.

DOQ Test Project

The DOQ test project is comprised of two phases. The first phase involves capturing the coordinates of certain TIGER feature intersection points. These points are called "anchor points" (AP). Once the anchor points have coordinates added from the DOQ they become "DOQ anchor points." The second phase includes the process of transforming all TIGER coordinates using the newly collected DOQ anchor point coordinate data.

This paper focuses on the first phase of the test project. Several tasks were accomplished to complete the first phase of the project, including:

- selecting a medium for use in data collection,
- selecting a test site,

- choosing a GIS,
- project set-up, preparing DOQ and TIGER/Line data, and
- anchor point data collection.

Selecting a Medium for Use in Data Collection

The GRaSS decided to use “images” as a second data collection technique (the first being GPS). The GRaSS chose DOQs as the type of image to use. A DOQ is a computer generated image of an aerial photograph in which displacements caused by camera orientation and terrain have been removed so that features are displayed in their true ground position. DOQs have the characteristics of a photograph with the capabilities of being used in a GIS. The reasons for the GRaSS selecting DOQs include their availability and ease of use. Other DOQ specifics that are significant to the project and notable are:

- Resolution - Resolution is the minimum distance between two adjacent features or the minimum size of a feature that can be detected by a remote sensing system. The ground (pixel) sample distance is approximately 1 meter. (The ground sample distance is the distance on the ground represented by each pixel in the X and Y components.)
- Scale of the Source Image - Approximately 1:40,000.
- Accuracy - DOQs must meet horizontal National Map Accuracy Standards (NMAS)ⁱⁱ at 1:24,000 and 1:12,000 scale, respectively. The NMAS specify that 90 percent of the well-defined points tested must fall within 40 feet (1/50 inch) at 1:24,000 scale and 33.3 feet (1/30 inch) at 1:12,000 scale. The vertical accuracy of the source Digital Elevation Model (DEM) must be equivalent or better than a Level 1 DEM, with a root-mean-square-error (RMSE) of no greater than 7.0 meters. The DOQ RMSE is the square root of the average of the squared discrepancies. These discrepancies are the differences in coordinate (X and Y) values derived by comparing the data tested with values determined during aerotriangulation or by an independent survey of higher accuracy.
- Projection/Datum - DOQs are cast on the Universal Transverse Mercator (UTM) projection on the North American Datum of 1983 (NAD83) with coordinates in meters.
- Geographic Extent - The geographic extent of a digital orthophoto quarter quad (DOQQ) is 3.75 minutes of latitude by 3.75 minutes of longitude, plus a minimum of 50 meters to a maximum of 300 meters of overedge included. One 7.5 x 7.5 minute DOQ is a mosaic of four DOQQs.
- File Size - Approximately 50 megabytes per DOQQ (uncompressed).
- Header - A header containing metadata is affixed to the beginning of each image file and is composed of numerous image characteristics. The information contained in the header is vital to DOQ set-up in ArcView.
- Availability - U.S. Geologic Survey (USGS) began production of DOQs in 1991 and the coterminous United States is expected to be complete by the year 2004. After completion, DOQs are expected to be updated on a ten year cycle for most areas and on a five year cycle for areas of rapid growth.
- Current Cost - Compressed DOQ files are distributed in a JPEG format on CD-ROM. Digital Orthophoto Quad - Quarter Quad CD-ROM black and white: handling charge (\$3.50), base charge (\$45.00), plus \$7.50 per file.

Selecting a Test Site

Two counties were selected by the GRaSS as test sites for both the DOQ and GPS test projects. Hampshire County, West Virginia and Newberry County, South Carolina fulfilled the criteria developed by GRaSS as being suitable for data collection. The following criteria were inclusive of both projects and were developed to facilitate the needs of both data collection methods:

- Rural County - The Census Bureau was concerned with collecting coordinate data from counties with populations that classify as rural. Local updates and source material is difficult

to acquire for many rural counties and it is often necessary for the Geography Division to conduct field work to obtain the information.

- Geography - Counties with differing geographic characteristics (extreme vs. moderate) provided for a more complete testing of methods. For example, in working with GPS technology, it was important to test its operability in both mountainous and plain type areas.
- DOQ Availability - In order to conduct the DOQ test project in any county, DOQs had to be available from USGS. It was preferred that DOQ data for counties be recently acquired as well.
- Staff Limitations - Limitations on the number of staff members and their time were considered when choosing a county for in-house and on-site fieldwork. Counties with limited road networks and that were in close proximity to the Census Bureau were preferred for time conservation and travel purposes.

Figure 1 lists the basic information and distinguishing characteristics of both Hampshire County, WV and Newberry County, SC.

	Hampshire County, West Virginia	Newberry County, South Carolina
State FIPS	54	45
County FIPS	027	071
1990 Population	16,498	33,172
Geographic Profile	Potomac highlands area of West Virginia	Central piedmont area of South Carolina
USGS DOQ Source Photographic Date (yyyymmdd)	19890317, 19890424, 19910411, and 19910417	19940201
Number of Quarter-Quad DOQs Providing Total County Coverage	61	64
Road Network Distance (miles measured in TIGER/Line '95 or '97)	1,332 (measured in TIGER/Line '95)	1,545 (measured in TIGER/Line '97)
1990 Housing Unit Count	10,168	13,777
State and Local Agencies Actively Supporting the GPS Test Project	2	6
Approximate Travel Time by Vehicle from Census Bureau Headquarters	3 hours	8 hours
Number of Anchor Points Created from TIGER/Line '97	3,591	3,722

Figure 1. Hampshire County, WV and Newberry County, SC Test Site Information.

Hampshire County was selected as the first test county. DOQ data collected for Hampshire County was limited to approximately 60 of the 3,591 existing TIGER anchor points. Limited DOQ data collection was due to the results of the corresponding GPS project for this county and not a limitation of the DOQ procedure. Data collected for Hampshire County using GPS technology contained errors and could not be used.

As a result, the remaining part of this paper discusses the procedure, obstacles encountered, and evaluation pertaining to the DOQ data collection for, the only test county, Newberry County, SC.

Why Choose ArcView?

ESRI's ArcView was chosen as the GIS software for this test project. The GRaSS believed that this software package was most conducive to the tasks that needed to be accomplished for the following four reasons:

- the tasks in this project that involved DOQ analysis, anchor point placement and file creation could be performed in ArcView,
- ArcView was capable of supporting and viewing image and feature formats that TIGER/Line '97 and DOQs could be converted into,
- ArcView was relatively easy to use and learn, and
- the U.S. Bureau of the Census, Geography Division had an ESRI site license.

Project Set-up

After selecting a medium for use in data collection (DOQs) and a test site (Newberry County, SC), the next step was to prepare the DOQ files, TIGER/Line '97 files, and the ArcView project for data collection.

DOQ Set-up Procedure

DOQs were stored as compressed images on a CD-ROM, making it necessary to decompress the DOQ data directly from the CD-ROM to the hard drive. The CD-ROM included a MS-DOS® executable to facilitate access to and use of text and image files. After decompressing and storing the images in the computer it was necessary to rename a DOQ file and change its original extension to a .bil extension. ArcView by default searched for .bil image files.

Each image file had an associated header file. A new header file was created using the same name as the newly named .bil image file. The new header file was given a .hdr extension and was in ASCII format. To create a new ASCII header file certain information contained in the original header file was required. The new header file contained the following eight lines of information:

Name	Definition
nrows	Number of rows or lines in image.
ncols	Number of columns or samples in image.
ulxmap	Upper left corner X value of pixel 1,1.
ulymap	Upper left corner Y value of pixel 1,1.
skipbytes	Number of bytes to skip that make up the header.
xdim	Dimension of pixel in X direction.
ydim	Dimension of pixel in Y direction.
nbands	Number of bands in image.

Figure 2. New Header File Requirements and Definitions.

One problem encountered while creating new header files for Newberry County was finding the appropriate value for *skipbytes*. It had been assumed that the value was fixed and would remain the same for all DOQ images. This was incorrect and affected the alignment of the DOQ image to other DOQ images and to the TIGER/Line file in the ArcView display. To initially resolve this problem, the correct information for the new ASCII header file was located manually; each item was determined separately. Once the problem was pinpointed, investigation uncovered an executable that was downloaded from the Internet and that read the original header file and organized each item according to its name.

After the image file was decompressed, renamed, and the new header file was created and named, the image was ready to be added as a theme to an ArcView project.

TIGER/Line Set-up Procedure

The TIGER/Line anchor point file, created by the GRaSS for the DOQ and GPS projects, is an ASCII text file containing four fields: 1) anchor point identification number (APID), 2) TIGER/Line longitude, 3) TIGER/Line latitude, and 4) anchor point quality rating. An anchor point is an intersection of three or more end nodes of TIGER/Line '97 Type 1 records, with only Type 1 roads, railroads and hydrographic features acceptable as an intersecting feature.

The quality of the anchor point was important to the overall project design, therefore, was recorded in the fourth field of the anchor point file. The GRaSS developed a quality-rating scheme from the existing source code in TIGER/Line '97. The source code represented the project or program from which the geographic entity and its properties originated.

The TIGER/Line reference file, also created by the GRaSS for the DOQ and GPS projects, is an ASCII text file extracted directly from TIGER/Line '97. The TIGER/Line reference file contains all Type 1 record line features. The purpose of this file was simply to serve as a further visual reference.

ArcView Project Set-up Procedure

In order to collect DOQ anchor point data, several components were required in the ArcView project. First, appropriate tables and themes were loaded into ArcView. A DOQ was loaded as an image theme. Both the TIGER/Line anchor point file and the TIGER/Line reference file were loaded into ArcView as tables. Once the TIGER/Line anchor point file was loaded, the information it contained was immediately added to a *View* as an *Event Theme*. Then the TIGER/Line reference file was loaded, and certain records were queried and extracted. By querying the Census Feature Class Code (CFCC) field of the TIGER/Line reference file, separate road, railroad, hydrographic, and boundary features could be loaded to the *View* as separate *Event Themes*. All TIGER anchor points and line features added to the *View* as *Event Themes* were able to be converted to shapefiles if desired. Finally, a new point theme was created, named "DOQ anchor point," and added to the *View*.

Data Collection

The GRaSS was interested in collecting only DOQ anchor points that represented intersecting roads, hydrographic features, and railroads that already existed in TIGER/Line '97. Therefore, the number of TIGER anchor points that existed in the TIGER/Line anchor point file equaled the number of new DOQ anchor points that were processed during data collection.

The GRaSS developed a simple procedure to create a new DOQ anchor point:

1. *Zoom-in* to a scale that the feature on both the DOQ and TIGER/Line can be identified.
2. Activate the TIGER anchor point theme.
3. Using the *Identify* Icon, click on the TIGER anchor point to display the *Identify Results* table. The TIGER anchor point's APID is displayed.
4. Adjust or *zoom-in* on the DOQ image to a comfortable level for DOQ anchor point placement.
5. Activate the DOQ anchor point theme and plot a new DOQ anchor point.
6. Open the associated table and choose *Table - Start Editing*.
7. Fill the appropriate fields with the attributes of the newly plotted DOQ anchor point.

For each DOQ anchor point created, five attributes were recorded in the associated ArcView *Attribute Table*:

1. APID - A sequential number given to the DOQ anchor point from the TIGER/Line anchor point file.
2. Rating - A rating scheme was developed by the GRaSS to assign each DOQ anchor point a grade that represented the quality of its placement. The following table shows the five ratings and their meaning.

Rating	Meaning
3	Excellent Placement

2	Fair
1	Unsure
0	Best Guess
-	Cannot be used

Figure 3. DOQ Anchor Point Rating.

Each rating was somewhat of a “guesstimate” by the analyst. Because of this, ratings varied among analysts.

- Reason - The GRaSS felt it was important to know the reason why analysts found some DOQ anchor points easier to place than others (and therefore assigned the DOQ anchor points different ratings). The GRaSS developed a classification scheme for identifying these reasons.

Problem Associated with DOQ		
D DOQ	R Road	U Unidentifiable
	H Hydrography	V V-type intersection
	I Intersection	S Star intersection
	L Railroad	P Pinpointing of intersection is difficult
		T Vegetation obscuring feature
		L Parallel features make issues
		B Blurs with background/other features
		H Shadow obscuring feature
		O Other
Problem Associated with TIGER/Line		
T TIGER/Line	R Road	A Additional AP not on DOQ
	H Hydrography	M Missing AP caused by merging AP
	I Intersection	L Lost road or railroad with random AP
	L Railroad	C Confusable with roads on DOQ
		T Topologically misplaced
		S Shape same as DOQ but displaced (>50m)
		N TIGER road not on nearby DOQ roads.
		D Digitized poorly
		R Scale is poor but shape is ok
		V V-type intersection – TIGER is jagged

Figure 4. DOQ Anchor Point Classification Scheme.

For example, the classification DHH denoted that on a DOQ the intersection of a stream and local road was difficult to plot due to a shadow obscuring the location of the stream. (The first D indicated the problem being associated with the DOQ, the first H indicated the stream as a hydrographic feature, and the final H indicated the specific problem, in this a case a shadow obscuring the stream.)

- Type - The TIGER/Line features that were intersecting at the TIGER anchor point were identified and recorded.

Intersecting Feature Identification	
H	Highway
L	Local Road
J	Jeep Road
R	Railroad
S	Stream

Figure 5. Intersection Feature Identification.

For example, if a local road and a stream create a four-way intersection, it is identified as L-L-S-S.

5. Interesting - DOQ anchor points that were particularly interesting or an exceptional case were noted with a "Y" for further investigation or to be used as examples. Otherwise, the field was left empty.

Analyst Thoughts

Determining whether an anchor point was easy or difficult to plot depended on the qualities of both the TIGER/Line feature and the DOQ image. GRaSS analysts discovered that DOQ anchor points were most easily plotted when:

- An intersection on a DOQ had defined edges, and the shade of the road, stream or railroad was not confused with other items in the nearby area.
- Vegetation cover did not obstruct features.
- The TIGER/Line road feature was in close proximity to the same feature on the DOQ.
- The source from which the TIGER/Line feature had been digitized was accurate and could be associated with a feature on the DOQ.

Analysts found it most challenging to plot DOQ anchor points when roads and streams were not identifiable due to the following reasons:

- Vegetation created a canopy over the road or intersection.
- Ground cover with little vegetation often appeared to be the same shade as a road.
- The digitized road in TIGER/Line did not appear to follow the shape of any of the roads within the nearby area on the DOQ.
- The middle of the star-type intersection was difficult to identify.
- Vegetation and hydrographic features were both displayed in a dark shade of gray and it was difficult to distinguish between the two.
- Often there were streams that appeared to be parallel to one another but there was only one stream that appeared in TIGER/Line.
- The digitized stream in TIGER/Line did not follow the shape of the stream in the DOQ.

From the above lists, it was apparent that most challenges and problems that occurred during data collection were associated with feature visibility and DOQ interpretation.

ArcView Performance

When project data collectors have little DOQ experience it becomes crucial to select a GIS that is easy to learn and use. More specifically, it is important to have image tools available in the software that cater to a minimum level of understanding.

To plot DOQ anchor points effectively and efficiently, features on a DOQ must be clear and distinguishable from other features. There are several functions in ArcView that facilitate DOQ anchor point placement on the image. Equally, there are several aspects lacking in ArcView that may have made DOQ anchor point placement easier or more accurate for a junior level geographer.

Successes Using ArcView

ArcView provides tools that allow the manipulation of a DOQ image. These tools lie primarily within the *Image Legend Editor*. The *Image Legend Editor* is prompted by activating the DOQ image theme and clicking on the *Edit Legend* button on the tool bar. The tools found most useful while analyzing DOQs are the image contrasting tools and zooming tools.

A gray scale image can be adjusted in the *Linear Lookup* option in the *Image Legend Editor*. The *Linear Lookup* provides a means to perform a contrast stretch, increase the contrast, soften the image, increase the brightness, and darken the image by modifying the graph that is provided. In some instances altering the image improved the visibility of particular features on the DOQ.

However, there was no particular aspect of the *Linear Lookup* that proved to be more beneficial than the other.

When analyzing a DOQ image and placing a DOQ anchor point, it is necessary to view the image at several different scales. The appropriate scale for an image to be at when placing DOQ anchor points varies depending on the particular DOQ image, the individual viewing the image, and the feature that is being analyzed. ArcView has several zooming tools that are helpful when placing anchor points such as *Zoom In*, *Zoom Out*, *Zoom to Previous*, *Zoom to Full Extent*, and *Zoom to Active Theme*.

Aside from image tools and DOQ analysis, ArcView shapefiles are documented and can be manipulated. For example, the GRaSS chose Microsoft Visual Basic® programming language to manipulate shapefiles by adding data and extracting data. The GRaSS was then able to load the manipulated files back into ArcView for visual interpretation.

Features Lacking in ArcView

Determining what tools are lacking in ArcView is difficult when an analyst has limited knowledge of photographic interpretation and GIS capabilities. How does an analyst decide what would benefit future DOQ projects without sufficient experience? If DOQs were adopted as the method to update TIGER at a national level, junior grade staff would be employed as primary operators. It is essential that the GIS software used for the project aid the analyst who is not an expert, nor wishes to be.

In an ArcView *View* window, zooming in and out of an image several times for the purpose of placing one anchor point can be time consuming. Each DOQ image is approximately 50 megabytes in size and it can take several seconds to redraw a simple image. Although several seconds does not seem significant, there are thousands of TIGER anchor points per county and the image can typically be redrawn five or more times per TIGER anchor point. In order to perform tasks in a timely manner, only one to two DOQ images can be active while placing DOQ anchor points. It would be helpful if ArcView were equipped to load large images at a faster pace.

Oftentimes, the clarity of features can vary within a single DOQ. It is necessary to manipulate the image several different ways to accommodate each DOQ anchor point that is being placed. When altering an image in ArcView, the entire image (one DOQQ) is affected. ArcView does not have the capability to adjust only one section of interest on the DOQ.

Overall, the ability to customize an ArcView project that caters specifically to gray scale image enhancement would be ideal. The above evaluation was written with the knowledge that a new ArcView Image Analysis Extension exists and contains features that may be of use to the DOQ project. However, the extension was not available to the GRaSS during the first phase of the DOQ test project and it is unclear that the extension will solve all the problems mentioned.

Conclusion

The analysts for the Newberry County, SC test project took approximately 200 hours to place 3,723 DOQ anchor points at a rate of approximately 20 DOQ anchor points per hour. After completing the first phase of the test project, the GRaSS believes DOQs are an efficient medium for use in data collection. However, the use of DOQs on a national level cannot be decided until the second phase of the project is complete, the results are validated, and all problems that have been encountered are addressed. Although the DOQ test project has thus far been successful, there are several components that need to be explored further. These components include potential software developments, image enhancement tools, and improving analyst interpretation skills.

ⁱ The use of brand names within this paper does not represent an endorsement of a company or its products.

ⁱⁱ The Federal Geographic Data Committee's endorsed 1998 National Standard for Spatial Data Accuracy (NSSDA) is intended to replace the 1947 National Map Accuracy Standard. The NSSDA was not available when the GRaSS was selecting the medium.

References

1997 TIGER/Line® Files Technical Documentation/prepared by the Bureau of the Census-Washington, DC, 1997.

Godwin, Leslie. (1998). Improving TIGER: A DOQ Test Project. In: Proceedings, GIS/LIS '98, Annual Conference & Exposition.

United States Geological Survey. Digital Orthophoto Program. URL:
<http://mapping.usgs.gov/www/ndop/>. 1998.

Environmental Systems Research Institute, Inc. ArcView GIS Version 3.1 Documentation. 1998.

Federal Geographic Data Committee. A Proposal for a National Spatial Data Infrastructure Standards Project. URL:
<http://www.fgdc.gov/standards/documents/proposals/progpas3.html>. 1999.

Kristen O'Grady
U.S. Bureau of the Census
Geography Division, GRaSS
GEO 7400
4700 Silver Hill Road
Washington, DC 20233-7400
Telephone: (301) 457-1056
E-mail: kogrady@geo.census.gov